

Fig 6c: Signature Tasks (Largest Performance Gaps)

Task	Terminus (%)	Codex (%)	Gap (%)	Favors
task_008	60.0	26.7	+33.3	Terminus
task_085	16.7	50.0	-33.3	Codex
task_037	53.3	23.3	+30.0	Terminus
task_043	63.3	33.3	+30.0	Terminus
task_010	56.7	33.3	+23.3	Terminus
task_016	50.0	73.3	-23.3	Codex
task_024	66.7	43.3	+23.3	Terminus
task_038	60.0	36.7	+23.3	Terminus
task_045	60.0	36.7	+23.3	Terminus
task_054	60.0	36.7	+23.3	Terminus
task_065	60.0	36.7	+23.3	Terminus
task_072	53.3	30.0	+23.3	Terminus
task_077	56.7	33.3	+23.3	Terminus
task_034	60.0	40.0	+20.0	Terminus
task_052	60.0	40.0	+20.0	Terminus
task_068	46.7	26.7	+20.0	Terminus
task_069	66.7	46.7	+20.0	Terminus
task_071	43.3	23.3	+20.0	Terminus
task_079	46.7	26.7	+20.0	Terminus
task_086	60.0	40.0	+20.0	Terminus